Review Problems

p. 685 #1,3,7,8,10,12,13

***1. Genetics***

Hypotheses:

H0: The distribution of traits is as specified by the ratio 1:3:3:9.

HA: The distribution of traits is not as specified.

Conditions:

1. Categorical data 1- Traits are categorical
2. SRS 2- Not stated as a random sample, so I will assume the sample is representative.
3. All exp. Counts > 5 3- All expected cell counts are > 5.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Trait | A; NC | A; C | F; NC | F; C |
| Observed | 10 | 22 | 31 | 59 |
| Expected | 7.625 | 22.875 | 22.875 | 68.625 |

All conditions have been met to use a Chi-Square distribution and a Chi-Square GOF test.

Mechanics:



Conclusion:

I fail to reject the null hypothesis b/c the P-Value of 0.1711 is greater than alpha = 0.05. We have sufficient evidence that the distribution of traits fits the ration of 1:3:3:9.

***3. Hard Water***

Hypotheses:

H0: *β*1 = 0 There is no association between calcium concentration in water and mortality rates for males.

HA: *β*1 ≠ 0 There is an association between calcium concentration in water and mortality rates for males.

Conditions:

Conditions are assumed to have been met.

Mechanics:



Conclusion:

I reject Ho b/c the P-Value of 7.74x10-9 is less than alpha = 0.05. There is statistically significant evidence that there is an association between calcium content in the water and mortality of males.

95% Confidence interval:



I am 95% confident that the slope of the population regression line between calcium content and mortality rate of males is between -2.27 and - 4.19 part per million per deaths per 100,000.

***OR***

I am 95% confident that for every increase of 1 part per million of calcium in the water there tends to be a decrease of between 2.27 to 4.19 deaths per 100,000.

***7. Back to Montana***

Hypotheses:

H0: Political party is independent of income level in Montana.

HA: There is an association between political party and income level in Montana.

Conditions:

1. Categorical data 1- political party and income level are categorical
2. SRS 2- Not stated as a random sample, so I will assume the sample is representative.
3. All exp. Counts > 5 3- All expected cell counts are > 5

Expected Counts (matrix B):

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  | Political Party | | |
|  |  | D | R | I |
| Income | Low | 24.119 | 22.396 | 14.485 |
| Middle | 30.772 | 28.574 | 14.653 |
| High | 29.109 | 27.03 | 13.861 |

All conditions have been met to use a Chi-Square distribution and a Chi-Square test of Independence.

Mechanics:



Conclusion:

We reject the Ho b/c the P-Value of 0.0018 is less than alpha = 0.05. There is statistically significant evidence that there is an association between political party and income in Montana.

***8. Wild Horses***

a) There were 38 herds studied.

b) Conditions:

1- SRS 1- assumed

2- linear data 2- scatterplot is linear with no outliers

3- independence 3- herds are independent of each other

4- normal residuals 4- normal probability plot of the residuals is straight

5- Equal variance 5- There is no apparent change of the spread of the residual plot.

All conditions have been met to use a t-distribution and a linear regression slope t-test.

c) 95% Confidence interval



I am 95% confident that the slope of the population regression line between adults and foals is between 0.131 to 0.177 horse/horse.

***OR***

I am 95% confident that for every increase of 1 adult horse in the herd there tends to be an increase of between 0.131 to 0.177 foals.

d) For every increase of 1 adult horse in the herd there tends to be an increase of 0.154 foals.

***e) ignore***

***10. AP Statistics Scores***

**a) Total School (Chi Square GOF test)**

Hypotheses:

H0: The distribution of scores for the school is the same as the national one.

HA: The distribution of scores for the school is not the same as the national one.

Conditions:

1. Categorical data 1- scores on the AP Exam are categorical
2. SRS 2- Not stated as a random sample, so I will assume the sample is representative.
3. All exp. Counts > 5 3- All expected cell counts are > 5

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Score | 5 | 4 | 3 | 2 | 1 |
| Observed | 26 | 36 | 19 | 10 | 6 |
| Expected | 11.155 | 22.698 | 24.153 | 18.527 | 20.467 |

All conditions have been met to use a Chi-Square distribution and a Chi-Square GOF test.

Mechanics:



Conclusion:

I reject Ho b/c the P-Value of 1.14x10-8 is less than alpha = 0.05. There is statistically significant evidence to say that the distribution of scores for the school is not the same as the national one.

**10)**

**b) Boys vs Girls (Chi-Square test for Homogeneity)**

***NOTE: Since the expected cell counts of the scores for 1 and 2 would be too small I combined the last two rows to get the following table:***

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Score | 5 | 4 | 3 | 2/1 |
| Boys | 13 | 21 | 6 | 11 |
| Girls | 13 | 15 | 13 | 5 |

Hypotheses:

H0: The distribution of scores on the AP Exam for boys is the same as for girls.

HA: The distribution of scores on the AP Exam for boys is different than for girls.

Conditions:

1. Categorical data 1- scores on the AP Exam and gender are both categorical
2. SRS 2- Not stated as a random sample, so I will assume the sample is representative.
3. All exp. Counts > 5 3- All expected cell counts are > 5

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Score | 5 | 4 | 3 | 2/1 |
| Boys | 13.67 | 18.928 | 9.9897 | 8.4124 |
| Girls | 12.33 | 17.072 | 9.0103 | 7.5876 |

All conditions have been met to use a Chi-Square distribution and a Chi-Square test of homogeneity.

Mechanics:



Conclusion:

I fail to reject Ho b/c the P-Value of 0.1336 is greater than alpha = 0.05. There is not enough evidence that an association exists between sex and score on the AP Exam for the high school.

***12. Twins***

Hypotheses:

H0: Duration of pregnancy and level of prenatal care are independent.

HA: Duration of pregnancy and level of prenatal care are dependent.

Conditions:

1. Categorical data 1- level of care and duration of pregnancy are both categorical
2. SRS 2- Not stated as a random sample, so I will assume the sample is representative.
3. All exp. Counts > 5 3- All expected cell counts are > 5

Expected counts (from Matrix B):

|  |  |  |  |
| --- | --- | --- | --- |
|  | Preterm (induced or cesarian) | Preterm (without procedures) | Term or postterm |
| Intensive | 16.676 | 15.579 | 28.745 |
| Adequate | 42.101 | 39.331 | 72.568 |
| Inadequate | 17.223 | 16.090 | 29.687 |

All conditions have been met to use a Chi-Square distribution and a Chi-Square test of Independence.

Mechanics:



Conclusion:

Since the P-Value is greater than alpha (0.1887 > 0.05) I fail to reject the null hypothesis. There is not enough evidence that an association exists between duration of pregnancy and level of prenatal care for twins.

***13. Twins, again***

Hypotheses:

H0: The distributions of pregnancy durations are the same for the three years.

HA: The distributions of pregnancy durations are different for the three years.

Conditions:

1. Categorical data 1- pregnancy duration and year are both categorical
2. SRS 2- Not stated as a random sample, so I will assume the sample is representative.
3. All exp. Counts > 5 3- All expected cell counts are > 5

Expected counts (Matrix B):

|  |  |  |  |
| --- | --- | --- | --- |
|  | 1990 | 1995 | 2000 |
| Preterm (induced or cesarian) | 12.676 | 13.173 | 17.150 |
| Preterm (without procedures) | 13.266 | 13.786 | 17.948 |
| Term or postterm | 25.058 | 26.04 | 33.902 |

All conditions have been met to use a Chi-Square distribution and a Chi-Square test of Homogeneity.

Mechanics:



Conclusion:

Since the P-Value is greater than alpha (0.9526 > 0.05) I fail to reject the null hypothesis. There is not enough evidence to suggest that the distribution of pregnancy durations was any different between the three years.